

ASSISTANT COMMISSIONER FOR PATENTS
Washington, D.C. 20231

#2
FORM PTO-1082
Case Docket No. 51270-245583
Date: October 27, 1998
Express Mail Label No. EM432436994US

Dear Sir:

Transmitted herewith for filing is the patent application of

Inventor(s): **Yasuo Yoshioka and Xavier Serra**

For: **CONVERTING APPARATUS OF VOICE SIGNAL BY MODULATION OF
FREQUENCIES AND AMPLITUDES OF SINUSOIDAL WAVE COMPONENTS**

jc525 U.S. PTO
09/18/021
10/27/98

Enclosed are:

- ☒ 6 Sheet(s) of drawings (___ informal)
- ☒ An assignment of the invention to Yamaha Corporation.
- ☐ An associate power of attorney.
- ☐ A verified statement to establish small entity status under 37 CFR 1.9 and 1.27.
- ☒ Declaration and Power of Attorney.
- ☒ Certified copy of Japan Patent Application No. 9-296050 filed October 28, 1997 from which priority is claimed under 35 USC §119.
- ☒ IDS enclosed. 1 with references.
- ☐ Preliminary Amendment

CALCULATION OF FEES							
ITEM		NO. OF CLAIMS FILED MINUS BASE*		NO. OF CLAIMS OVER BASE	X SM/LG ENTITY FEE	\$ AMOUNT	\$ FEE
A	TOTAL CLAIMS FEE	27	- 20* =	7	X \$11 or X \$22	\$154	
B	INDEPENDENT CLAIMS FEE**	6	- 3* =	3	X \$41 or X 82	\$246	
C	SUBTOTAL - ADDITIONAL CLAIMS FEE (ADD FINAL COLUMN IN LINES A + B)						\$400
D	MULTIPLE-DEPENDENT CLAIMS FEE				SMALL ENTITY FEE = \$135 LARGE ENTITY FEE = \$270		\$
E	BASIC FEE*				SMALL ENTITY FEE = \$395 LARGE ENTITY FEE = \$790		\$790
F	TOTAL FILING FEE (ADD TOTALS FOR LINES C, D, AND E)						\$1,190
F	ASSIGNMENT RECORDING FEE					\$ 40	\$40
	**LIST INDEPENDENT CLAIMS						

Please charge my Deposit Account No. \$0
16-1805 the amount of

**A copy of this letter is
enclosed.**

☒ A check in the amount of \$1,190

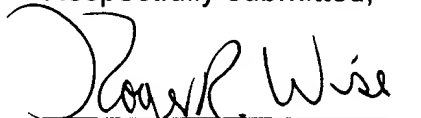
to cover the filing fee is
enclosed.

☒ A check in the amount of \$40

to cover Assignment
Recordation fee is enclosed.

- ☒ The Commissioner is hereby authorized to charge payment of the following fees associated with this communication or credit any overpayment to Deposit Account No. 16-1805.
- ☒ Any additional filing fees required under 37 CFR 1.16.
- ☒ Any patent application processing fees under 37 CFR 1.17.
- ☐ The Commissioner is hereby authorized to charge payment of the following fees during the pendency of this application or credit any overpayment to Deposit Account No. 16-1805.
- ☐ Any patent application processing fees under 37 CFR 1.17.
- ☐ The issue fee set in 37 CFR 1.18 at or before mailing of the Notice of Allowance, pursuant to 37 CFR 1.311(b).
- ☐ Any filing fees under 37 CFR 1.16 for presentation of extra claims.

Respectfully submitted,



Roger R. Wise
Reg. No. 31,204

Date: October 27, 1998

PILLSBURY MADISON & SUTRO LLP
725 South Figueroa Street, Suite 1200
Los Angeles, CA 90017-5443
Telephone: (213) 488-7100
Facsimile: (213) 629-1033

日 本 国 特 許 庁

PATENT OFFICE
JAPANESE GOVERNMENT

JCS25 U.S. PTO
09/181021
10/27/98

別紙添付の書類に記載されている事項は下記の出願書類に記載されて
いる事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed
with this Office.

出 願 年 月 日

Date of Application:

1997年10月28日

出 願 番 号

Application Number:

平成 9年特許願第296050号

出 願 人
Applicant(s):

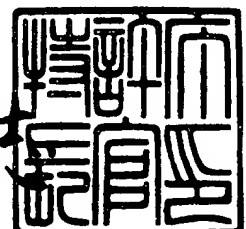
ヤマハ株式会社

CERTIFIED COPY OF
PRIORITY DOCUMENT

1998年10月 2日

特許庁長官
Commissioner,
Patent Office

山 建 彦



【書類名】 特許願

【整理番号】 C26812

【提出日】 平成 9年10月28日

【あて先】 特許庁長官殿

【国際特許分類】 G10L 5/00

【発明の名称】 音声変換装置

【請求項の数】 9

【発明者】

 【住所又は居所】 静岡県浜松市中沢町10番1号 ヤマハ株式会社内

 【氏名】 吉岡 靖雄

【特許出願人】

 【識別番号】 000004075

 【氏名又は名称】 ヤマハ株式会社

【代理人】

 【識別番号】 100098084

 【弁理士】

 【氏名又は名称】 川▲崎▼ 研二

【選任した代理人】

 【識別番号】 100104798

 【弁理士】

 【氏名又は名称】 山下 智典

【手数料の表示】

 【予納台帳番号】 038265

 【納付金額】 21,000円

【提出物件の目録】

 【物件名】 明細書 1

 【物件名】 図面 1

 【物件名】 要約書 1

 【包括委任状番号】 9600966

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 音声変換装置

【特許請求の範囲】

【請求項1】 入力された音声信号から複数の正弦波成分を抽出する正弦波成分抽出手段と、

参照用音声のピッチ情報を記憶した参照ピッチ情報記憶手段と、

前記参照ピッチ情報記憶手段から読み出されたピッチ情報に基づいて前記各正弦波成分の周波数を調整する周波数調整手段と、

前記周波数調整手段によって周波数が調整された後の前記各正弦波成分を合成して合成波形を生成する合成波形生成手段と

を具備することを特徴とする音声変換装置。

【請求項2】 入力された音声信号から複数の正弦波成分を抽出する正弦波成分抽出手段と、

参照用音声から抽出された複数の正弦波成分の振幅を示す振幅情報を記憶する振幅情報記憶手段と、

振幅情報記憶手段から読み出された振幅情報に基づいて前記正弦波成分の振幅を調整する振幅調整手段と、

前記振幅調整手段によって振幅が調整された後の前記各正弦波成分を合成して合成波形を生成する合成波形生成手段と

を具備することを特徴とする音声変換装置。

【請求項3】 入力された音声信号から複数の正弦波成分を抽出する正弦波成分抽出手段と、

参照用音声のピッチ情報を記憶した参照ピッチ情報記憶手段と、

前記参照用音声から抽出された複数の正弦波成分の振幅を示す振幅情報を記憶する振幅情報記憶手段と、

振幅情報記憶手段から読み出された振幅情報に基づいて前記正弦波成分の振幅を調整する振幅調整手段と、

前記参照ピッチ情報記憶手段から読み出されたピッチ情報に基づいて前記各正弦波成分の周波数を調整する周波数調整手段と、

前記周波数調整手段および前記振幅調整手段によって周波数および振幅が調整された後の前記各正弦波成分を合成して合成波形を生成する合成波形生成手段とを具備することを特徴とする音声変換装置。

【請求項4】 前記周波数調整手段は、前記正弦波成分に対する前記ピッチ情報の反映の度合いを所定のパラメータに応じて変化させることを特徴とする請求項1または3いずれかに記載の音声変換装置。

【請求項5】 前記参照ピッチ記憶手段は、音階の単位で変化する音階的ピッチと、前記音階的ピッチに対するピッチの揺らぎを示す揺らぎ成分とを記憶し、前記周波数調整手段は、前記音階的ピッチと前記揺らぎ成分との双方に基づいて前記正弦波成分の周波数を調整することを特徴とする請求項1、3または4記載の音声変換装置。

【請求項6】 前記振幅調整手段は、前記正弦波成分に対する前記振幅情報の反映の度合いを所定のパラメータに応じて変化させることを特徴とする請求項2または3記載の音声変換装置。

【請求項7】 前記参照音声の音量変化を示す音量情報を記憶する音量情報記憶手段と、前記音量情報記憶手段から読み出される音量情報に基づいて、前記合成波形の音量を調整する音量調整手段をさらに具備することを特徴とする請求項1乃至6いずれかに記載の音声変換装置。

【請求項8】 前記入力された音声信号中のピッチの有無を判定するピッチ判定手段と、前記ピッチ判定手段がピッチ無しの判定をした場合に、前記合成波形に変えて前記入力された音声信号を出力する切換手段を具備することを特徴とする請求項1乃至7いずれかに記載の音声変換装置。

【請求項9】 前記正弦波成分抽出手段が抽出した正弦波成分と前記入力された音声信号との残差成分を求める残差成分抽出手段と、

前記残差成分抽出手段が抽出した残差成分を前記合成波形に加える加算手段とをさらに具備することを特徴とする請求項1乃至8いずれかに記載の音声変換装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

この発明は、処理対象となる音声を、目標とする他の音声に近似させる音声変換装置に関する。

【0002】

【従来の技術】

入力された音声の周波数特性などを変えて出力する音声変換装置は種々開発されており、例えば、カラオケ装置の中には、歌手の歌った歌声のピッチを変換して、男性の声を女性の声に、あるいはその逆に変換させるものもある（例えば、特表平8-508581号）。

【0003】

【発明が解決しようとする課題】

しかしながら、従来の音声変換装置においては、音声の変換は行われるものの、単に声質を変えるだけに止まっていたので、例えば、誰かの声に似せるように変換するということはできなかった。

また、声質だけでなく、歌い方までも誰かに似させるという、ものまねのような機能があれば、カラオケ装置などにおいては大変に面白いが、従来の音声変換装置ではこのような処理は不可能であった。

【0004】

この発明は、上述した事情に鑑みてなされたもので、声質を目標とする声に似させることができる音声変換装置を提供することを目的としている。

また、入力された歌手の音声を、目標とする人の歌い方に似せることができる音声変換装置を提供することを目的とする。

【0005】

【課題を解決するための手段】

この発明は、上記課題を解決するために、請求項1に記載の音声変換装置においては、入力された音声信号から複数の正弦波成分を抽出する正弦波成分抽出手段と、参照用音声のピッチ情報を記憶した参照ピッチ情報記憶手段と、前記参照

ピッチ情報記憶手段から読み出されたピッチ情報に基づいて前記各正弦波成分の周波数を調整する周波数調整手段と、前記周波数調整手段によって周波数が調整された後の前記各正弦波成分を合成して合成波形を生成する合成波形生成手段とを具備する。

【0006】

また、請求項2に記載の音声変換装置においては、入力された音声信号から複数の正弦波成分を抽出する正弦波成分抽出手段と、参照用音声から抽出された複数の正弦波成分の振幅を示す振幅情報を記憶する振幅情報記憶手段と、振幅情報記憶手段から読み出された振幅情報に基づいて前記正弦波成分の振幅を調整する振幅調整手段と、前記振幅調整手段によって振幅が調整された後の前記各正弦波成分を合成して合成波形を生成する合成波形生成手段とを具備することを特徴とする。

【0007】

また、請求項3に記載の音声変換装置においては、入力された音声信号から複数の正弦波成分を抽出する正弦波成分抽出手段と、参照用音声のピッチ情報を記憶した参照ピッチ情報記憶手段と、前記参照用音声から抽出された複数の正弦波成分の振幅を示す振幅情報を記憶する振幅情報記憶手段と、振幅情報記憶手段から読み出された振幅情報に基づいて前記正弦波成分の振幅を調整する振幅調整手段と、前記参照ピッチ情報記憶手段から読み出されたピッチ情報に基づいて前記各正弦波成分の周波数を調整する周波数調整手段と、前記周波数調整手段および前記振幅調整手段によって周波数および振幅が調整された後の前記各正弦波成分を合成して合成波形を生成する合成波形生成手段とを具備することを特徴とする。

【0008】

また、請求項4に記載の音声変換装置にあっては、請求項1または3に記載の音声変換装置にあって、前記周波数調整手段は、前記正弦波成分に対する前記ピッチ情報の反映の度合いを所定のパラメータに応じて変化させることを特徴とする。

【0009】

また、請求項5に記載の音声変換装置にあっては、請求項1、3または4記載の音声変換装置において、前記参照ピッチ記憶手段は、音階の単位で変化する音階的ピッチと、前記音階的ピッチに対するピッチの揺らぎを示す揺らぎ成分とを記憶し、前記周波数調整手段は、前記音階的ピッチと前記揺らぎ成分との双方に基づいて前記正弦波成分の周波数を調整することを特徴とする。

【0010】

また、請求項6に記載の音声変換装置にあっては、請求項2または3記載の音声変換装置において、前記振幅調整手段は、前記正弦波成分に対する前記振幅情報の反映の度合いを所定のパラメータに応じて変化させることを特徴とする。

【0011】

また、請求項7に記載の音声変換装置にあっては、請求項1乃至6いずれかに記載の音声変換装置において、前記参照音声の音量変化を示す音量情報を記憶する音量情報記憶手段と、前記音量情報記憶手段から読み出される音量情報に基づいて、前記合成波形の音量を調整する音量調整手段をさらに具備することを特徴とする。

【0012】

また、請求項8に記載の音声変換装置においては、請求項1乃至7いずれかに記載の音声変換装置において、前記入力された音声信号中のピッチの有無を判定するピッチ判定手段と、前記ピッチ判定手段がピッチ無しの判定をした場合に、前記合成波形に変えて前記入力された音声信号を出力する切換手段を具備することを特徴とする。

【0013】

また、請求項9に記載の音声変換装置にあっては、請求項1乃至8いずれかに記載の音声変換装置において、前記正弦波成分抽出手段が抽出した正弦波成分と前記入力された音声信号との残差成分を求める残差成分抽出手段と、前記残差成分抽出手段が抽出した残差成分を前記合成波形に加える加算手段とをさらに具備することを特徴とする。

【0014】

【発明の実施の形態】

1. 第1実施形態の基本構成

次に、本発明の実施の形態について説明する。図1は、この発明の第1実施形態の構成を示すブロック図である。なお、この実施例は、この発明による音声変換装置をカラオケ装置に適用し、ものまねを行うことができるカラオケ装置を構成した例である。

【0015】

始めに、この実施例の原理について説明する。まず、ものまねの対象となる人の歌を分析し、そのピッチおよび正弦波成分の振幅を記憶しておく。そして、歌い手の音声から正弦波成分を抽出し、この正弦成分に対して、ものまねの対象の人のピッチと正弦波成分の振幅を反映させる。そして、反映させた正弦波成分を合成して合成波形を作成し、これを増幅して出力する。また、この際に反映させる度合いを所定のパラメータで調整できるようにする。

以上の処理により、ものまねの対象となる人の声質や歌い方が反映された音声波形が作成され、これがカラオケ演奏とともに出力される。

【0016】

2. 第1実施形態の詳細構成

図1において、1はマイクであり、歌い手の声を収拾し、その音声信号 S_v を出力する。この音声信号 S_v は、高速フーリエ変換部2によって解析処理され、その周波数スペクトルが検出される。高速フーリエ変換部2の処理は、所定のフレーム単位で行われるため、周波数スペクトルは各フレーム毎に順次作成される。ここで、音声信号 S_v とフレームとの関係を図2に示す。図2に示す記号 FL がフレームであり、この実施形態においては前のフレーム FL と一部重なるように設定されている。

【0017】

次に、3は周波数スペクトルのピークを検出するピーク検出部である。例えば、図3に示すような周波数スペクトルに対して、×印を付けたピーク値を検出する。このピーク値は、周波数値と振幅値の座標として (F_0, A_0) 、 $(F_1,$

A1)、(F2、A2)……(FN、AN)というように各フレームについて一組にして出力される。ここで、図2に各フレームに対応するピーク値の組を模式的に示す。

次に、ピーク検出部3から出力された各フレームについてのピーク値の組は、ピーク連携部4において、前後のフレームについて連携が判断され、連携すると認められるピーク値については、データ列となるように連携処理される。ここで、この連携処理について、図4を参照して説明する。

今、図4の部分(A)に示すようなピーク値が前のフレームにおいて検出され、同図の部分(B)に示すようなピーク値が次のフレームにおいて検出されたとする。この場合、ピーク連携部4は、前のフレームで検出された各ピーク値(F0、A0)、(F1、A1)、(F2、A2)……(FN、AN)に対応するピーク値が今回のフレームでも検出されたか否かを調べる。対応するピーク値があるか否かの判断は、前のフレームで検出されたピーク値の周波数を中心にした所定範囲内に今回のピークを検出されるか否かによって行われる。図4の例では、ピーク値(F0、A0)、(F1、A1)、(F2、A2)……については、対応するピーク値が発見されているが、ピーク値(FK、AK)については、対応するピーク値は発見されていない。

【0018】

ピーク連携部4は、対応するピーク値が発見された場合は、それらを時系列順に繋げて一組のデータ列として出力する。なお、対応するピーク値が発見されない場合は、そのフレームについての対応ピークは無しということを示すデータに置き換える。ここで、図5は、ピーク周波数F0とF1の変化の一例を示している。このような変化が振幅A0、A1、A2……についても同様に発生する。この場合、ピーク連携部4から出力されるデータ列は、フレームの間隔おきに出力される離散的な値である。

なお、ピーク連携部4から出力されるピーク値を、以後において、確定成分という。これは、元の信号(すなわち、音声信号Sv)のうち正弦波の要素として確定的に置き換えられる成分という意味である。また、置き換えられた各正弦波(厳密には、正弦波のパラメータである振幅と周波数)の各々については、部分

成分と呼ぶことにする。

【0019】

次に、補間・波形発生部5は、ピーク連携部4から出力される確定成分について補間処理を行い、補間後の確定成分に基づいた波形を発生を行う。この場合の補間のピッチは、最終出力信号（後述するアンプ50に入力される直前の信号）のサンプリングレート（例えば、44.1 KHz）に対応したピッチで行われる。前述した図5に示す実線は、ピーク値のF0、F1に対して補間処理が行われた場合のイメージを示している。

ここで、補間・波形発生部5の構成を図7に示す。この図に示す5a、5a…は、各々部分波形発生部であり、指示された周波数値および振幅値に応じた正弦波を発生する。ただし、本実施例における部分成分（F0、A0）、（F1、A1）、（F2、A2）…は、各々補間のピッチに従って時事刻々変化していくものであるから、部分波形発生部5a、5a…から出力される波形は、その変化に従った波形になる。すなわち、ピーク連携部4からは部分成分（F0、A0）、（F1、A1）、（F2、A2）…が順次出力され、その各々について補間処理が行われるから、各部分波形発生部5a、5a…は、所定の周波数領域内で周波数と振幅が変動する波形を出力する。そして、各部分波形発生部5a、5a…から出力された波形は、加算部5bにおいて加算合成される。したがって、補間・波形発生部5の出力信号は、元信号（すなわち音声信号Sv）から確定成分を抽出した波形になる。

【0020】

次に、図1に示す偏差検出部6は、補間・波形発生部5から出力された確定成分波形と音声信号Svとの偏差を検出する。この偏差成分を、以後においては残差成分Srdという。この残差成分は、音声に含まれる無声成分を多く含む。一方、前述の確定成分は有声成分に対応するものである。ところで、誰かの声に似せるには、有声音についてだけ処理を行い、無声音については処理はあまり必要がない。そこで、この実施形態においては、有声母音成分に対応する確定成分について音声変換処理を行うようにしている。次に、図1に示す10は分離部であり、ピーク連携部4が出力するデータ列の中から周波数値F0～FNと振幅値A

0～ANとを分離する。ピッチ検出部11は、分離部10から供給される周波数値に基づいて各フレーム毎のピッチを検出する。この場合のピッチ検出は、分離部10が出力する周波数値のうち最も低い値から所定数（例えば3個程度）の周波数値を選択し、それらの周波数値を所定の重み付けをした後に、それらの平均を算出してピッチPSとする。また、ピッチ検出部11は、ピッチを検出することができないフレームについては、ピッチ無しを示す信号を出力する。ピッチ無しのフレームとは、そのフレーム内の音声信号Svがほとんど無声音やノイズによって構成されている場合である。このようなフレームについては、周波数スペクトルが倍音構成とならないので、ピッチ無しと判定する。

【0021】

次に、20は音声を似せようとする対象（以下、ターゲットという）の情報が記憶されているターゲット情報記憶部である。ターゲット情報記憶部20は、曲毎にターゲットの情報を記憶している。ターゲットの情報は、ターゲットの音声の音階的なピッチを抽出したピッチ情報PToと、ピッチの揺らぎ成分PTfと、確定的な振幅成分（分離部10が出力する振幅値A0、A1、A2……と同種の成分）とを有しており、これらの情報は、音階的ピッチ記憶部21、ゆらぎピッチ記憶部22および確定的振幅成分記憶部23に各々記憶されている。

ターゲット情報記憶部20は、カラオケ演奏に同期して、上述した各情報を読み出すようになっている。カラオケ演奏は、図1に示す演奏部27において行われる。演奏部27は、カラオケ用の曲データを予め記憶しており、図示せぬ選択手段によって選択された曲データを楽曲の進行順に読み出してアンプ50に供給する。このとき、演奏部27は、楽曲名とその進行状況を示す制御信号Scをターゲット情報記憶部20に供給し、ターゲット情報記憶部20は、制御信号に基づいて上述した各情報を読み出していく。

【0022】

次に、音階的ピッチ記憶部21から読み出されたピッチ情報PToは、割合制御部30においてピッチPSと混合される。この場合の混合は、次の式に基づいて行われる。

$$(1.0-\alpha)*PS+\alpha*PTo \quad \cdots \cdots (1)$$

ここで、 α は0から1までの値をとるパラメータであり、割合制御部30から出力される信号は、 $\alpha=0$ でピッチPSに等しくなり、 $\alpha=1$ でピッチ情報PT α に等しくなる。また、パラメータ α は、操作者がパラメータ設定部25を操作することによって任意の値が設定される。パラメータ設定部25においては、後述するパラメータ β 、 γ も設定可能になっている。

【0023】

次に、図1に示すピッチ正規化部12は、分離部10から出力される各周波数値F0～FNをピッチPSで割り、周波数値を正規化する。正規化された各周波数値F0/PS～FN/PS（ディメンジョンは無名数）は、乗算部15によって割合制御部からの信号と乗算され、そのディメンジョンは再び周波数となる。この場合、パラメータ α の値により、マイク1から音声を入力している歌手（以下、シンガーという）のピッチの影響が強くなるか、あるいは、ターゲットのピッチの影響が強くなるかが決定される。

割合制御部31は、ゆらぎピッチ記憶部22から出力される揺らぎ成分PTfにパラメータ β （ $0 \leq \beta \leq 1$ ）を乗算して乗算部16に出力する。この場合、揺らぎ成分PTfは、セントの単位でピッチ情報PT α に対する偏差を示している。従って、割合制御部31においては、揺らぎ成分PTfを1200（1オクターブは1200セント）で除し、それに対し2のべきをとる演算を行う。すなわち、以下の演算を行う。

$$\text{POW}(2, (\text{PTf} * \beta / 1200))$$

この演算結果と乗算部15の出力信号が乗算され、さらに、乗算部16の出力信号は、乗算部17において、トランスポーズ制御部32の出力信号と乗算される。トランスポーズ制御部32は、移調を行う音程に応じた値を出力するものである。どの程度の移調を行うかは、任意に設定されるが、通常は、移調なしが設定されるか、あるいは、オクターブ単位の変化が指定される。オクターブ単位の変化が指定されるのは、ターゲットが男性でシンガーが女性（あるいはその逆）の場合のように、歌う音程にオクターブの差がある場合などのときである。

以上のようにして、ピッチ正規化部12から出力された周波数値は、ターゲットのピッチ、揺らぎ成分が付与され、さらに、必要であればオクターブ変換が行

われた後に混合部40に入力される。

【0024】

次に、図1に示す13は、振幅検出部であり、分離部10から供給される振幅値 A_0 、 A_1 、 A_2 ……の平均値 MS をフレーム毎に検出する。振幅正規化16においては、振幅値 A_0 、 A_1 、 A_2 ……をその平均値で割り、振幅値を正規化する。割合制御部18においては、確定的振幅成分記憶部23から読み出される確定的振幅成分 AT_0 、 AT_1 、 AT_2 ……（これらは正規化されている）と正規化された振幅値とを混合する。混合の度合いはパラメータ γ に従って行われる。確定的振幅成分 AT_0 、 AT_1 、 AT_2 ……を AT_n （ $n=1, 2, 3$ ……）で表し、振幅正規化部16から出力される振幅値を AS_n' （ $n=1, 2, 3$ ……）で表すと、割合制御部18の動作は次の演算で表される。

$$(1-\gamma)*AS_n' + \gamma*AT_n$$

γ はパラメータ設定部25において適宜設定されるパラメータであり、0から1までの値をとる。 γ が大きいほど、ターゲットの影響を強く受ける。音声信号の正弦波成分の振幅は、声質を決めるものであるから、 γ が大きいほどターゲットの声質に近くなる。

割合制御部18の出力信号は、乗算部19において、平均値 MS と乗算される。すなわち、正規化された信号から振幅を直接表す信号に変換される。

【0025】

次に、混合部40においては、振幅値と周波数値が混合される。この混合された信号は、シンガーの音声信号 S_v の確定成分にターゲットの確定成分が加味されたものとなる。なお、パラメータ α 、 β 、 γ の値によっては、ターゲット側100%の確定成分となる。

この確定成分（正弦波である部分成分の集合）は、補間・波形発生部41に供給される。補間・波形発生部41は前述した補間・波形発生部5（図7参照）と同様に構成されており、混合部40から出力される確定成分に含まれる部分成分を補間し、補間後の各部分成分に基づいて部分波形を発生し、それらを合成する。合成された波形は、加算部42において残差成分 S_{rd} と加算され、切換部43を介してアンプ50に供給される。切換部43は、ピッチ検出部11がピッチ

を検出できないフレームについては、加算部42が出力する合成された信号に換えてシンガーの音声信号 S_v をアンプ50に供給する。これはノイズや無声音については、上述した各種処理を行う必要がないので、元の信号を直接出力した方がよいためである。

【0026】

3. 第1実施形態の動作

次に、上記構成によるこの実施形態の動作について説明する。まず、曲が指定されると、演奏部27において当該曲の曲データが読み出され、これに基づく楽音信号が形成されてアンプ50に供給される。そして、シンガーは、その伴奏にのって歌を歌い出す。これにより、マイク1から音声信号 S_v が出力され、この音声信号 S_v の確定成分がピーク検出部3によってフレーム毎に順次抽出される。例えば、図6の部分(1)のような抽出結果が得られる(なお、図6は1つのフレームにおいて得られる信号を示す)。そして、部分成分についてフレーム毎の連携が付けられ、これが分離部10において分離されて周波数値と振幅値に分けられて、図6の部分(2)、(3)に示すようになる。さらに、周波数値はピッチ正規化部12によって正規化され、図6に示す部分(4)に示すようになる。振幅値も同様に正規化され、図6の部分(5)に示すようになる。図6の部分(5)に示す正規化された振幅値に対して、部分(6)に示すようなターゲットの正規化された振幅値が混合され、部分(8)に示すような振幅値となる。この混合の割合はパラメータ γ によって決定される。

【0027】

一方、図6の部分(4)に示す周波数値に対しては、ターゲットのピッチ情報 $P T o$ および揺らぎ成分 $R T f$ が混合され、部分(7)に示すような周波数値となる。この混合の割合は、パラメータ α 、 β によって決定される。そして、図6の部分(7)、(8)に示すような周波数値と振幅値が混合部40によって混合され、同図の部分(9)に示すような新たな確定成分が得られる。この新たな確定成分は、補間・波形発生部41によって合成波形となり、残差成分 $S r d$ と混合された後にアンプ50に出力される。

以上の結果、カラオケの伴奏とともに、シンガーの歌が出力されるが、その声

質および歌い方などは、ターゲットの影響を大きく受け、パラメータ α 、 β 、 γ の値を1にすると、ターゲットそのものの声質および歌い方となる。このようにして、あたかもターゲットの物まねをしているような歌が出力される。

【0028】

4. 変形例

(1) 図7に示すように、ターゲットの音声の音量の変化を示す正規化音量データを記憶する正規化音量データ記憶部60を設けてもよい。この正規化音量データ記憶部60から読み出した正規化音量データに対し、乗算部61においてパラメータ k と乗算した後、切換手段43から出力され合成波形と乗算部62において乗算する。以上のような構成によれば、ターゲットの歌の抑揚についても模写することができる。この場合の模写の度合いは、パラメータの値によって決定される。したがって、反映させてい程度に応じてパラメータ k の値を設定すればよい。

【0029】

(2) 対象とするフレームにピッチがあるか無いかの検出は、本実施形態においては、ピッチ検出部11において行ったが、ピッチ有無の判定は、これに限らず、例えば、音声信号 S_v の状態から直接判定してもよい。

(3) 正弦波の抽出は、この実施形態で用いた方法に限らない。要は、音声信号に含まれる正弦波を抽出できればよい。

(4) 本実施形態においては、ターゲットのピッチや確定的振幅成分を記憶したが、これに換えて、ターゲットの音声そのものを記憶し、それを読み出してリアルタイム処理によってピッチと確定的振幅成分を抽出してもよい。すなわち、本実施形態でシンガーの声に対して行った処理と同様の処理をターゲットの音声に対して行ってもよい。

(5) 本実施形態においては、ターゲットの音階的ピッチと揺らぎ成分の双方を処理に用いたが、音階的ピッチだけを用いてもよい。また、音階的ピッチと揺らぎ成分を混合したピッチデータを作成し、これを用いるようにしてもよい。

(6) 本実施形態においては、シンガーの音声信号の確定成分（正弦波成分の集合）の周波数と振幅の双方を変換したが、いずれか一方だけを行うようにしても

よい。

(7) 本実施形態においては、補間・波形発生部 5、41 について、発振器を用いるいわゆるオシレーター方式を採用したが、これに限らず、例えば、逆 FFT を用いてもよい。

【0030】

【発明の効果】

以上説明したように、この発明によれば、ターゲットの声質や歌い方に似せるようにして音声を変換することができる。

【図面の簡単な説明】

【図 1】 この発明の一実施形態の構成を示すブロック図である。

【図 2】 同実施形態におけるフレームの状態を示す図である。

【図 3】 同実施例における周波数スペクトルのピーク検出を説明するための説明図である。

【図 4】 同実施形態におけるフレーム毎のピーク値の連携を示す図である。

【図 5】 同実施形態における周波数値の変化状態を示す図である。

【図 6】 同実施形態における処理過程における確定成分の変化状態を示すグラフである。

【図 7】 同実施形態における補間・波形発生部 5、41 の構成を示すブロック図である。

【図 8】 同実施形態における変形例の構成を示すブロック図である。

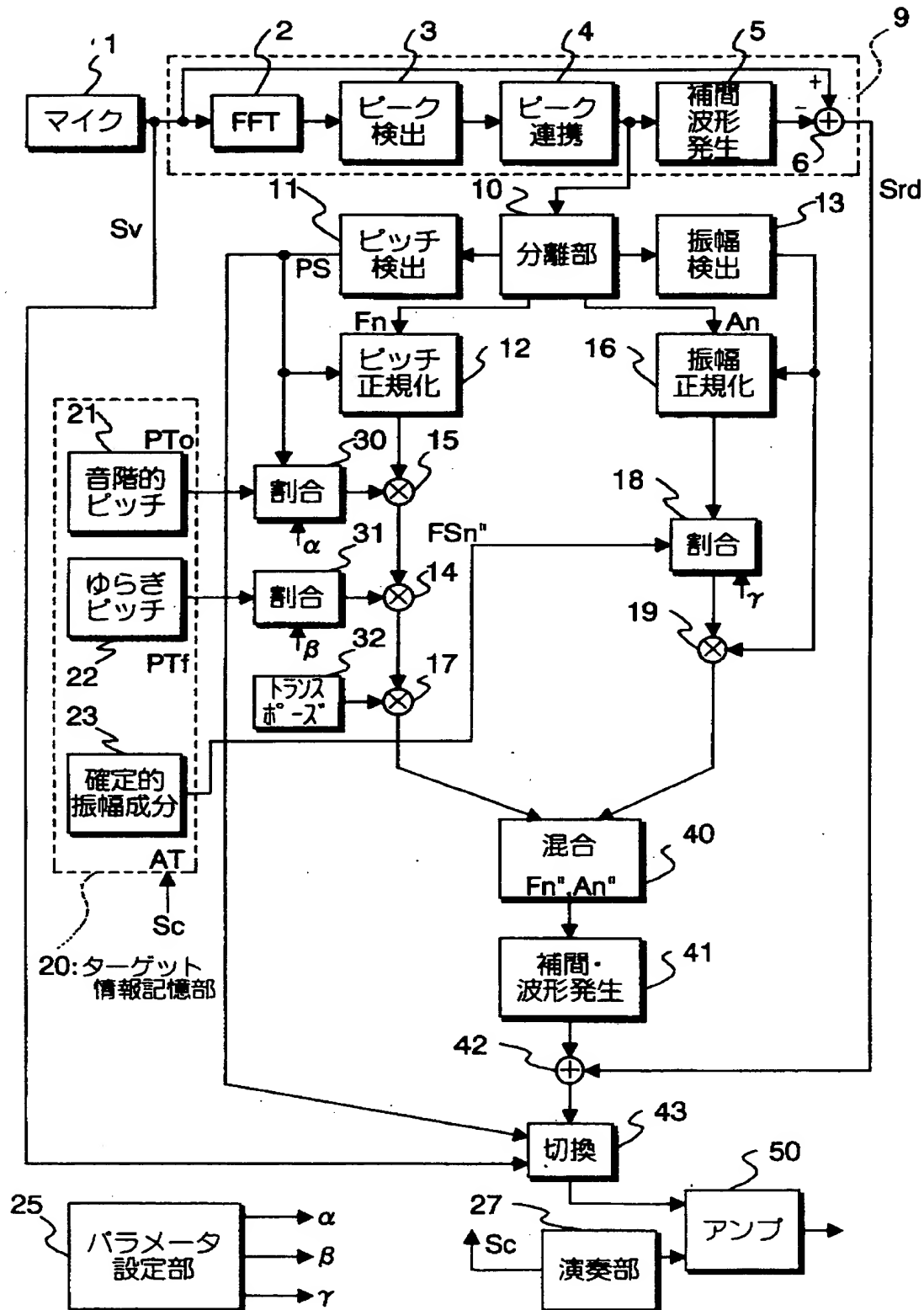
【符号の説明】

2 ……高速フーリエ変換部（正弦波成分抽出部）、3 ……ピーク検出部（正弦波成分抽出部）、4 ……ピーク連携部（正弦波成分抽出部）、5 ……補間波形発生部（残差成分抽出手段）、6 ……偏差検出部（残差成分抽出手段）、11 ……ピッチ検出部（ピッチ判定手段）、12 ……ピッチ正規化部、13 ……振幅検出部、14、15、17 ……乗算部（周波数調整手段）、16 ……振幅正規化部、17 ……振幅検出部、20 ……ターゲット情報記憶部（参照ピッチ記憶手段、振幅情報記憶手段）、25 ……パラメータ設定部、30、31 ……割合制御部（周

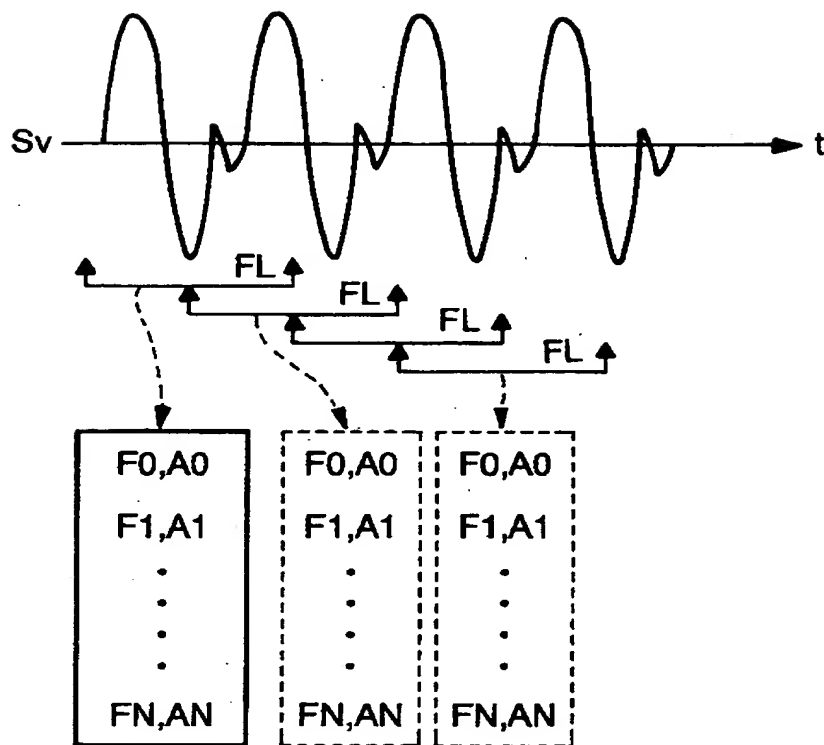
波数調整手段)、40……混合部(合成波形発生手段)、41……補間・波形発生部(合成波形発生手段)、42……加算部、43……切換部(切換手段)、60……正規化音量データ記憶部(音量情報記憶手段)、61、62……乗算部(音量調整手段)。

【書類名】 図面

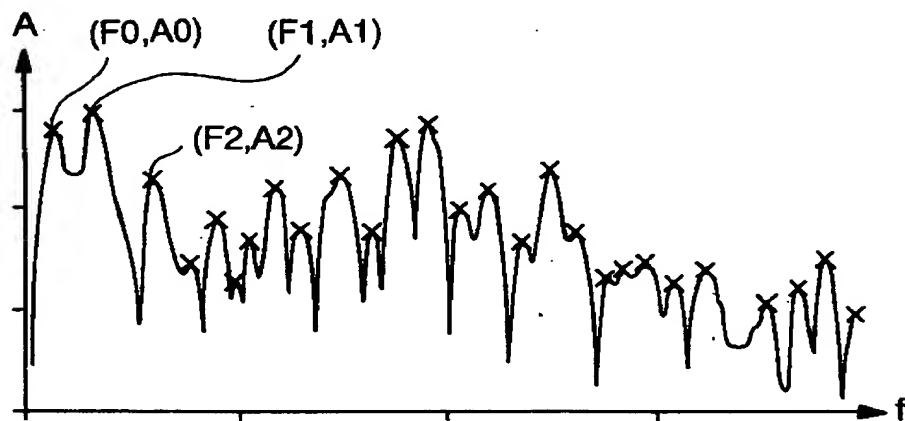
【図1】



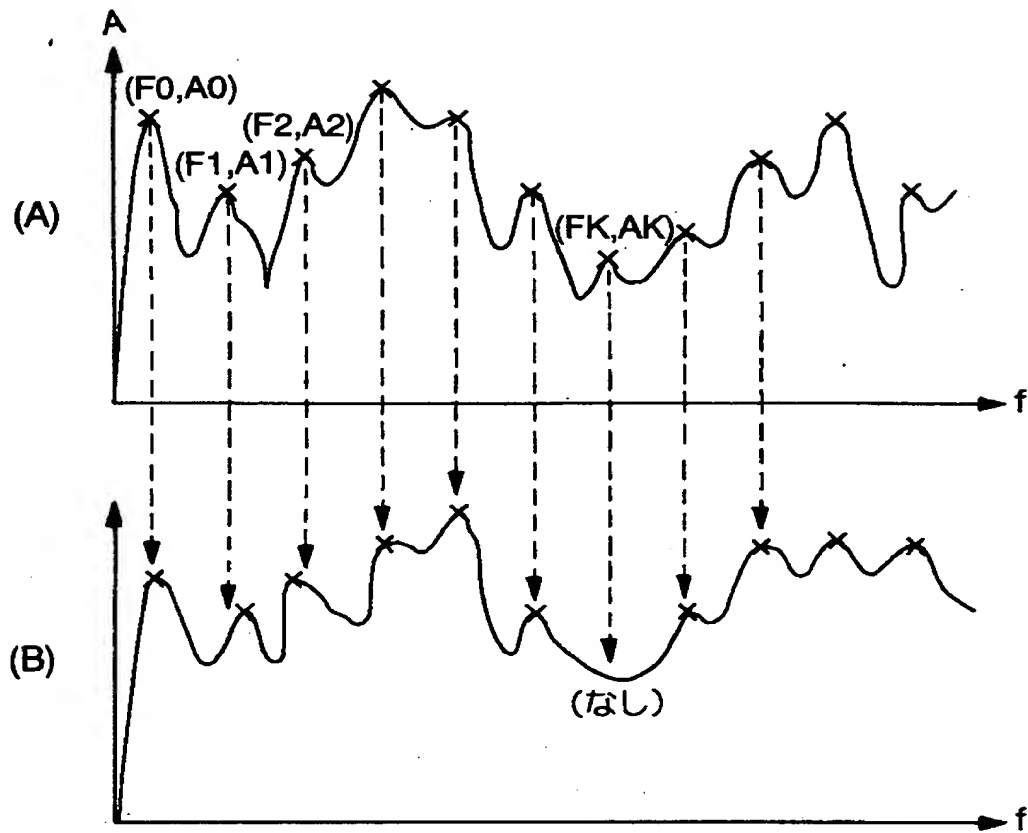
【图2】



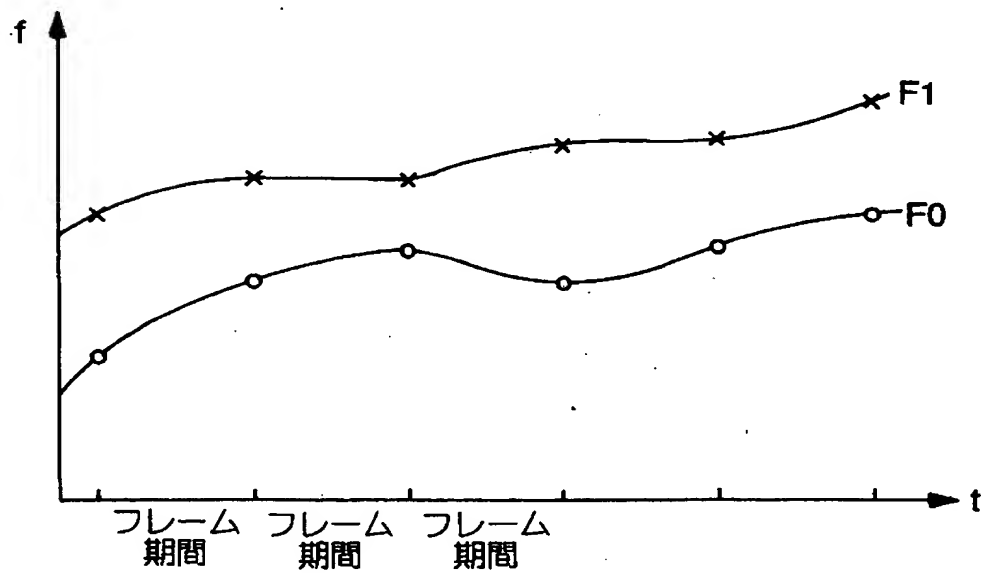
【图3】



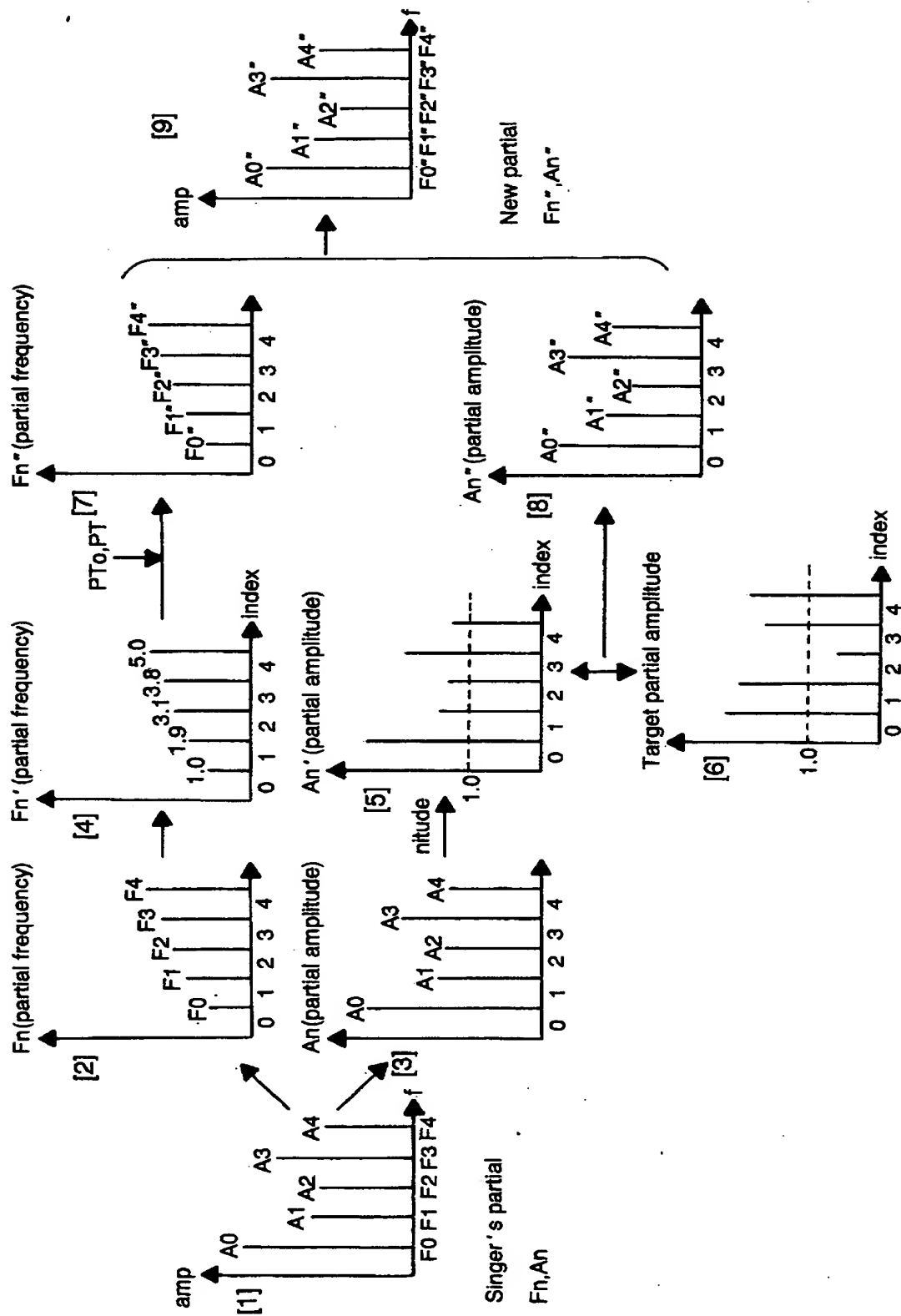
【図4】



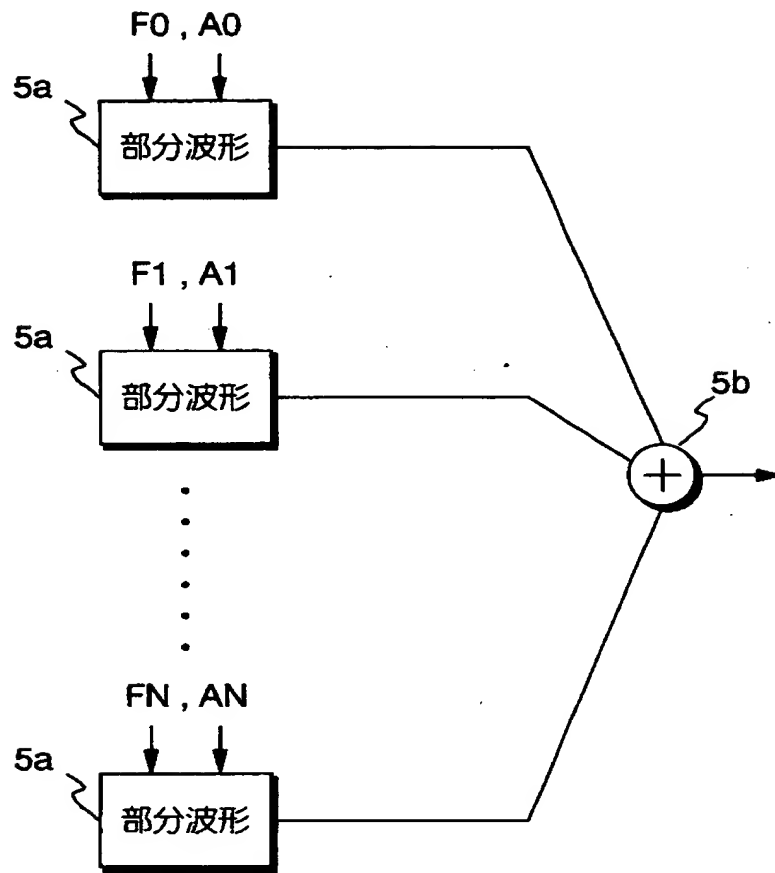
【図5】



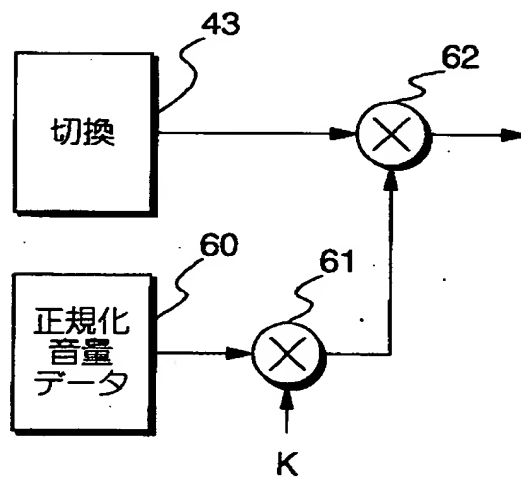
【图6】



【図7】



【図8】



【書類名】 要約書

【要約】

【課題】 入力された音声を目的とする人に似た音声に変換する。

【解決手段】 まず、ものまねの対象となる人の歌を分析し、そのピッチおよび正弦波成分の振幅をターゲット情報記憶部20に記憶しておく。そして、歌手の音声信号 S_v から正弦波成分を抽出し、この正弦成分に対して、ものまねの対象の人のピッチ P_{T0} と正弦波成分の振幅 A_T を反映させる。そして、反映させた正弦波成分を混合部40合成し、補間・波形発生部42で合成波形を作成し、これを増幅して出力する。一方、演奏部27では曲データに基づいてカラオケ伴奏信号生成し、アンプ50に出力する。以上の処理により、ものまねの対象となる人の声質や歌い方が反映された音声波形が作成され、これがカラオケ演奏とともに出力される。

【選択図】 図1

【書類名】 職権訂正データ
【訂正書類】 特許願

<認定情報・付加情報>

【特許出願人】

【識別番号】 000004075
【住所又は居所】 静岡県浜松市中沢町10番1号
【氏名又は名称】 ヤマハ株式会社

【代理人】 申請人

【識別番号】 100098084
【住所又は居所】 東京都中央区日本橋三丁目2番16号 八重洲マ
ヤビル5階 朝日特許事務所

【氏名又は名称】 川▲崎▼ 研二

【選任した代理人】

【識別番号】 100104798
【住所又は居所】 東京都中央区日本橋三丁目2番16号 八重洲マ
ヤビル5階 朝日特許事務所

【氏名又は名称】 山下 智典

【書類名】 手続補正書

【提出日】 平成10年 5月12日

【あて先】 特許庁長官 殿

【事件の表示】

【出願番号】 平成 9年特許願第296050号

【補正をする者】

【事件との関係】 特許出願人

【識別番号】 000004075

【氏名又は名称】 ヤマハ株式会社

【代理人】

【識別番号】 100098084

【弁理士】

【氏名又は名称】 川▲崎▼ 研二

【手続補正 1】

【補正対象書類名】 特許願

【補正対象項目名】 発明者

【補正方法】 変更

【補正の内容】

【発明者】

【住所又は居所】 静岡県浜松市中沢町10番1号 ヤマハ株式会社内

【氏名】 吉岡 靖雄

【発明者】

【住所又は居所】 スペイン バルセロナ カルデデュー08440 2-
2 ビスカイア19

【氏名】 ザビエル セラ

【手続補正 2】

【補正対象書類名】 明細書

【補正対象項目名】 0020

【補正方法】 変更

- 【補正の内容】 1
- 【手続補正 3】
- 【補正対象書類名】 明細書
- 【補正対象項目名】 0023
- 【補正方法】 変更
- 【補正の内容】 2
- 【手続補正 4】
- 【補正対象書類名】 明細書
- 【補正対象項目名】 0027
- 【補正方法】 変更
- 【補正の内容】 3
- 【手続補正 5】
- 【補正対象書類名】 明細書
- 【補正対象項目名】 0028
- 【補正方法】 変更
- 【補正の内容】 4
- 【手続補正 6】
- 【補正対象書類名】 明細書
- 【補正対象項目名】 0029
- 【補正方法】 変更
- 【補正の内容】 5
- 【手続補正 7】
- 【補正対象書類名】 明細書
- 【補正対象項目名】 符号の説明
- 【補正方法】 変更
- 【補正の内容】 6
- 【手続補正 8】
- 【補正対象書類名】 要約書
- 【補正対象項目名】 全文

【補正方法】 変更

【補正の内容】 7

【提出物件の目録】

【物件名】 宣誓書 2

【物件名】 宣誓書訳文 1

【物件名】 譲渡書 1

【物件名】 譲渡書訳文 1

【物件名】 理由書 1

【0020】

次に、図1に示す偏差検出部6は、補間・波形発生部5から出力された確定成分波形と音声信号 S_v との偏差を検出する。この偏差成分を、以後においては残差成分 S_{rd} という。この残差成分は、音声に含まれる無声成分を多く含む。一方、前述の確定成分は有声成分に対応するものである。ところで、誰かの声に似せるには、有声音についてだけ処理を行い、無声音については処理はあまり必要がない。そこで、この実施形態においては、有声成分に対応する確定成分について音声変換処理を行うようにしている。次に、図1に示す10は分離部であり、ピーク連携部4が出力するデータ列の中から周波数値 $F_0 \sim F_N$ と振幅値 $A_0 \sim A_N$ とを分離する。ピッチ検出部11は、分離部10から供給される周波数値に基づいて各フレーム毎のピッチを検出する。この場合のピッチ検出は、たとえば、分離部10が出力する周波数値のうち最も低い値から所定数（例えば3個程度）の周波数値を選択し、それらの周波数値を所定の重み付けをした後に、それらの平均を算出してピッチ PS とする。また、ピッチ検出部11は、ピッチを検出することができないフレームについては、ピッチ無しを示す信号を出力する。ピッチ無しのフレームとは、そのフレーム内の音声信号 S_v がほとんど無声音やノイズによって構成されている場合である。このようなフレームについては、周波数スペクトルが倍音構成とならないので、ピッチ無しと判定する。

【0023】

次に、図1に示すピッチ正規化部12は、分離部10から出力される各周波数値 $F_0 \sim F_N$ をピッチ PS で割り、周波数値を正規化する。正規化された各周波数値 $F_0/PS \sim F_N/PS$ （ディメンジョンは無名数）は、乗算部15によって割合制御部からの信号と乗算され、そのディメンジョンは再び周波数となる。この場合、パラメータ α の値により、マイク1から音声を入力している歌手（以下、シンガーという）のピッチの影響が強くなるか、あるいは、ターゲットのピッチの影響が強くなるかが決定される。

割合制御部31は、ゆらぎピッチ記憶部22から出力される揺らぎ成分 PTf にパラメータ β （ $0 \leq \beta \leq 1$ ）を乗算して乗算部16に出力する。この場合、揺らぎ成分 PTf は、セントの単位でピッチ情報 PTo に対する偏差を示している。従って、割合制御部31においては、揺らぎ成分 PTf を1200（1オクターブは1200セント）で除し、それに対し2のべきをとる演算を行う。すなわち、以下の演算を行う。

$$POW(2, (PTf * \beta / 1200))$$

この演算結果と乗算部15の出力信号が乗算され、さらに、乗算部14の出力信号は、乗算部17において、トランスポーズ制御部32の出力信号と乗算される。トランスポーズ制御部32は、移調を行う音程に応じた値を出力するものである。どの程度の移調を行うかは、任意に設定されるが、通常は、移調なしが設定されるか、あるいは、オクターブ単位の変化が指定される。オクターブ単位の変化が指定されるのは、ターゲットが男性でシンガーが女性（あるいはその逆）の場合のように、歌う音程にオクターブの差がある場合などのときである。

以上のようにして、ピッチ正規化部12から出力された周波数値は、ターゲットのピッチ、揺らぎ成分が付与され、さらに、必要であればオクターブ変換が行われた後に混合部40に入力される。

【0027】

一方、図6の部分(4)に示す周波数値に対しては、ターゲットのピッチ情報 $P T o$ および揺らぎ成分 $P T f$ が混合され、部分(7)に示すような周波数値となる。この混合の割合は、パラメータ α 、 β によって決定される。そして、図6の部分(7)、(8)に示すような周波数値と振幅値が混合部40によって混合され、同図の部分(9)に示すような新たな確定成分が得られる。この新たな確定成分は、補間・波形発生部41によって合成波形となり、残差成分 $S r d$ と混合された後にアンプ50に出力される。

以上の結果、カラオケの伴奏とともに、シンガーの歌が出力されるが、その声質および歌い方などは、ターゲットの影響を大きく受け、パラメータ α 、 β 、 γ の値を1にすると、ターゲットそのものの声質および歌い方となる。このようにして、あたかもターゲットの物まねをしているような歌が出力される。

【0028】

4. 変形例

(1) 図8に示すように、ターゲットの音声の音量の変化を示す正規化音量データを記憶する正規化音量データ記憶部60を設けてもよい。この正規化音量データ記憶部60から読み出した正規化音量データに対し、乗算部61においてパラメータ k と乗算した後、切換手段43から出力され合成波形と乗算部62において乗算する。以上のような構成によれば、ターゲットの歌の抑揚についても模写することができる。この場合の模写の度合いは、パラメータの値によって決定される。したがって、反映させたい程度に応じてパラメータ k の値を設定すればよい。

【0029】

(2) 対象とするフレームにピッチがあるか無いかの検出は、本実施形態においては、ピッチ検出部11において行ったが、ピッチ有無の判定は、これに限らず、例えば、音声信号Svの状態から直接判定してもよい。

(3) 正弦波成分の抽出は、この実施形態で用いた方法に限らない。要は、音声信号に含まれる正弦波成分を抽出できればよい。

(4) 本実施形態においては、ターゲットのピッチや確定的振幅成分を記憶したが、これに換えて、ターゲットの音声そのものを記憶し、それを読み出してリアルタイム処理によってピッチと確定的振幅成分を抽出してもよい。すなわち、本実施形態でシンガーの声に対して行った処理と同様の処理をターゲットの音声に対して行ってもよい。

(5) 本実施形態においては、ターゲットの音階的ピッチと揺らぎ成分の双方を処理に用いたが、音階的ピッチだけを用いてもよい。また、音階的ピッチと揺らぎ成分を混合したピッチデータを作成し、これを用いるようにしてもよい。

(6) 本実施形態においては、シンガーの音声信号の確定成分（正弦波成分の集合）の周波数と振幅の双方を変換したが、いずれか一方だけを行うようにしてもよい。

(7) 本実施形態においては、補間・波形発生部5、41について、発振器を用いるいわゆるオシレーター方式を採用したが、これに限らず、例えば、逆FFTを用いてもよい。

【符号の説明】

2 ……高速フーリエ変換部（正弦波成分抽出部）、3 ……ピーク検出部（正弦波成分抽出部）、4 ……ピーク連携部（正弦波成分抽出部）、5 ……補間波形発生部（残差成分抽出手段）、6 ……偏差検出部（残差成分抽出手段）、11 ……ピッチ検出部（ピッチ判定手段）、12 ……ピッチ正規化部、13 ……振幅検出部、14、15、17 ……乗算部（周波数調整手段）、16 ……振幅正規化部、20 ……ターゲット情報記憶部（参照ピッチ記憶手段、振幅情報記憶手段）、25 ……パラメータ設定部、30、31 ……割合制御部（周波数調整手段）、40 ……混合部（合成波形発生手段）、41 ……補間・波形発生部（合成波形発生手段）、42 ……加算部、43 ……切換部（切換手段）、60 ……正規化音量データ記憶部（音量情報記憶手段）、61、62 ……乗算部（音量調整手段）。

【書類名】 要約書

【要約】

【課題】 入力された音声を目的とする人に似た音声に変換する。

【解決手段】 まず、ものまねの対象となる人の歌を分析し、そのピッチおよび正弦波成分の振幅をターゲット情報記憶部20に記憶しておく。そして、歌い手の音声信号 S_v から正弦波成分を抽出し、この正弦成分に対して、ものまねの対象の人のピッチ P_{T0} と正弦波成分の振幅 A_T を反映させる。そして、反映させた正弦波成分を混合部40で合成し、補間・波形発生部42で合成波形を作成し、これを増幅して出力する。一方、演奏部27では曲データに基づいてカラオケ伴奏信号生成し、アンプ50に出力する。以上の処理により、ものまねの対象となる人の声質や歌い方が反映された音声波形が作成され、これがカラオケ演奏とともに出力される。

【選択図】 図1

29808900435



宣誓書

平成 年 月 日

住所 静岡県浜松市中沢町10番1号

ヤマハ株式会社内

氏名 吉岡 靖雄



下記発明は ザビエル セラ との共同発明であることに相違ありません。

記

1. 出願番号

平成9年特許願第296050号

2. 発明の名称

音声変換装置

DECLARATION

I, Xavier Serra, of Biscaia 19, 2-2, 08440 Cardedeu, Barcelona, Spain do
hereby sincerely and solemnly declare that I invented the following invention
with Yasuo Yoshioka:

Japanese Patent Application No. 9-296050

"Voice Changing Apparatus"

This 22nd day of April, 1998

By: 

Xavier Serra

29808900435



宣 誓 書 (訳 文)

スペイン バルセロナ カルデデュ-08440 2-2 ビスカイア19に居住
の私、ザビエル セラは、次の発明を 吉岡 靖雄氏と発明したことに相違ないこと
をここに宣誓します。

日本特許出願 平成9年第296050号
音声変換装置

1998年4月22日
ザビエル セラ (署名)

29808900435



DEED OF ASSIGNMENT

Assignees,

Name : YAMAHA CORPORATION

Address: 10-1, Nakazawa-cho, Hamamatsu-shi, Shizuoka-ken, Japan

"Voice Changing Apparatus"

I, the undersigned assignor, Xavier Serra,
do hereby affirm that I have assigned my rights to obtain a patent of the
above-mentioned invention to the above-identified assignees.

This 22nd day of April, 1997

Assignor,

Address: Biscaia 19, 2-2, 08440 Cardedeu, Barcelona, Spain

Name: Xavier Serra

(Signature)

29808900435



譲渡書(訳文)

譲受人

名前: ヤマハ株式会社

住所: 日本国 静岡県 浜松市 中沢町 10-1

音声変換装置

下記署名の譲渡人である私、ザビエル セラは、ここに、上記の発明の特許を受ける権利を 上記譲受人に譲渡したことを確認します。

1997年4月22日

譲渡人

住所: スペイン, バルセロナ, カルデデュ-08440, 2-2, ビスカイア19

名前: ザビエル セラ

(署名) ザビエル セラ

29808900435



理 由 書

1. 事件の表示

平成9年特許願第296050号

2. 発明の名称

音声変換装置

本件は、発明者2名で出願すべきところを 発明者「ザビエル セラ」を記載し忘れて出願してしまいました。

この誤記に至ったのは、代理人の不注意によるものであります。つきましては、上記発明者の追加致したく、ここに必要書類を添えて提出致しますので、手続きを宜しくお願い申し上げます。

以上

平成10年5月12日

出願人代理人

川崎 研二

【書類名】 職権訂正データ
【訂正書類】 手続補正書

<認定情報・付加情報>

【補正をする者】
【識別番号】 000004075
【住所又は居所】 静岡県浜松市中沢町10番1号
【氏名又は名称】 ヤマハ株式会社
【代理人】 申請人
【識別番号】 100098084
【住所又は居所】 東京都中央区日本橋三丁目2番16号 八重洲マ
ヤビル5階 朝日特許事務所
【氏名又は名称】 川▲崎▼ 研二
【提出された物件の記事】
【提出物件名】 宣誓書 1
宣誓書訳文 1
譲渡書 1
譲渡書訳文 1
理由書 1

出 願 人 履 歴 情 報

識別番号 [000004075]

1. 変更年月日	1990年 8月22日
[変更理由]	新規登録
住 所	静岡県浜松市中沢町10番1号
氏 名	ヤマハ株式会社